Supplementary Materials for

"Fast spread of COVID-19 in Europe and the US and its implications: even modest public health goals require comprehensive intervention"

Ruian Ke[1,*], Steven Sanche[1,2], Ethan Romero-Severson[1], Nick Hengartner[1]

Affiliations:

[1]T-6 Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, NM87544, USA.

[2]T-CNLS Center for Nonlinear Studies, Los Alamos National Laboratory, NM87544, USA.

[*]Correspondences to: Ruian Ke (rke@lanl.gov)

Phone: 1-505-667-7135

Mail: Mail Stop K710,

T-6 Theoretical Biology and Biophysics,

Los Alamos National Laboratory,

NM87544, USA.

Supplementary Text

1. Data Collection

We extracted data from https://github.com/CSSEGISandData/COVID-19 (John Hopkins Center for Systems Science and Engineering). The date of data access and extraction is March 31, 2020. The data consists of time series of case confirmations and deaths by country. We used data for the following countries: France (FR), Italy (IT), Spain (SP), Germany (GR), Belgium (BE), Switzerland (SW), Netherlands (NT), United Kingdom (UK) and the US (US). We calculated daily case confirmation and death incidence from cumulative counts.

A few entries in the collected data show signs of bulk reporting. The following procedure was performed to aggregate data where bulk reporting was suspected. It is applied to both case confirmation and death incidence data:

- If an increase of more than 1000% was observed between two consecutive dates, the data was aggregated over the two-day period and accordingly, incidence is considered to be over the two days.
- If the number of reported cases was zero on a date, followed by a case incidence of 10 or more, the data was aggregated over the two-day period.
- If two consecutive days had zero incidence, followed by an incidence of 20 or more, the data was aggregated over the three-day period.

2. Mathematical model

Standard SEIR model for COVID-19

We first construct a basic susceptible (S)- exposed (E) – infected (I) – recovered (R) model for COVID-19 and then extend upon the basic model. The ordinary differential equations are:

$$\frac{dS}{dt} = -\beta \frac{S}{N} I$$
$$\frac{dE}{dt} = \beta \frac{S}{N} I - kE$$
$$\frac{dI}{dt} = kE - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

where N is the total number of population, β is the infectivity, $1/k$ is the latent period, i.e. from infection to onset of infectiousness, $1/\gamma$ is the infectious period. Note that for implicitly, we assumed that R compartment include both recovered and dead individuals in this basic model and γ is the overall rate of individuals leaving I compartment.

We are interested in the dynamics of early exponential growth. We make the common assumption of a constant susceptible population ($S(t)=N$) during this period, and get a reduced EIR model:

$$\frac{dE}{dt} = \beta I - kE$$
$$\frac{dI}{dt} = kE - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

The long-term solution of the EIR model is driven by a single exponential whose rate is determined by the dominant eigenvalue, λ, of the Jacobian matrix of the EIR model. The exponential growth rate r, is thus the same as the dominant eigenvalue, λ:

$$r = \lambda = \frac{\sqrt{(k+\gamma)^2 + 4k(\beta - \gamma)} - (k+\gamma)}{2}$$

For a specific value of the exponential growth rate r, we can calculate β as:

$$\beta = \frac{1}{k} r^2 + \frac{k+\gamma}{k} r + \gamma$$

If we let $I^*(t)$ be the total number of infected individuals, and define that $I^*(t) = E(t) + I(t) = I_0^* e^{rt}$, we get the following expressions for $E(t)$ and $I(t)$:

$$I(t) = \frac{k+r}{k+r+\beta} I_0^* e^{rt}$$
$$E(t) = \frac{\beta}{k+r+\beta} I_0^* e^{rt}$$

we calculate the true daily incidence predicted by the model as:

$$\Omega(t) = \int_{t-1}^{t} \beta I(s)\, ds = \int_{t-1}^{t} \beta \frac{k+r}{k+r+\beta} I_0^* e^{rs}\, ds = \frac{\beta(k+r)}{r(k+r+\beta)} I_0^* \left(e^{rt} - e^{r(t-1)}\right)$$

Extending the model to consider case confirmation

We extend the EIR model to consider case confirmation. In this model, we consider that among newly infected individuals at time *t*, βI(t), a fraction, $\theta(t)$, of them will be tested at a later time, and it takes $1/g$ period of time to get case confirmation. The ODE model is then:

$$\frac{dE}{dt} = \beta I - kE$$
$$\frac{dI}{dt} = kE - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$
$$\frac{dC_0}{dt} = \theta(t)\beta I - gC_0$$
$$\frac{dC_1}{dt} = gC_0$$

where C_0 and C_1 correspond to the individuals who is tested during infection and the cumulative confirmed cases, respectively.

We are interested in new confirmed cases during a day, and thus it is reasonable to assume $\theta(t)$ does not change (at θ_t) in one day period. With this assumption, we solve the ODE above, we get:

$$C_0(t) = \theta_t \frac{1}{(g+r)} \frac{\beta(k+r)}{k+r+\beta} I_0^* e^{rt}$$

The daily new confirmed case count, $\Psi(t)$, is then:

$$\Psi(t) = \int_{t-1}^{t} gC_0(s)\, ds = \theta_t \frac{g}{(g+r)} \frac{\beta(k+r)}{r(k+r+\beta)} I_0^* \left(e^{rt} - e^{r(t-1)}\right)$$
$$= \theta_t \frac{g}{r(g+r)} \Omega(t)$$

Extending the EIR model to explicitly consider death

Now, we extend the EIR model to explicitly consider death of infected individuals. Let *X* be the case fatality ratio, and we assume an Erlang distribution for the period between onset of infectiousness to death. We get the following model:

$$\frac{dI_{d,1}}{dt} = XkE - ndI_{d,1}$$
$$\cdots$$
$$\frac{dI_{d,n}}{dt} = nd(I_{d,n-1} - I_{d,n})$$
$$\frac{dD}{dt} = ndI_{d,k}$$

where *n* is the shape parameter of the Erlang distribution for the period from onset of infectiousness to death, and $1/d$ is the mean duration from the beginning of infectious period to death. Note that we assumed that the infected individuals who eventually die have the same latent period for simplicity. The sensitivity of our conclusions to this assumption is tested in sensitivity analysis where we varied the range of the duration from onset of infectiousness to death.

Solving the ODEs above, we get:

$$I_{d,i}(t) = \frac{(nd)^{i-1}}{(r+nd)^i} X \frac{k\beta}{k+r+\beta} I_0^* e^{rt}, \;\; i = 1..n$$

$$D(t) = (\frac{nd}{r+nd})^n X \frac{k\beta}{r(k+r+\beta)} I_0^* e^{rt}$$

Then, the daily death count, $\Phi(t)$, is:

$$\Phi(t) = \int_{t-1}^{t} ndI_{d,n}\, ds = \left(\frac{nd}{r+nd}\right)^n X \frac{\beta k}{r(k+r+\beta)} I_0^* \left(e^{rt} - e^{r(t-1)}\right)$$

$$= \left(\frac{nd}{r+nd}\right)^n X \frac{k}{k+r}\, \Omega(t)$$

3. Discussion of parameter values and ranges

The latent period, 1/k

We set 1/*k*=3 days as a baseline. Previously we and others estimated the incubation period, i.e. the duration between infection and symptom onset to be between 4-6 days[1-3]. It is possible that infectiousness onset starts 1-2 days before symptom onset[4]. Therefore, we set the latent period at 3 days with a range of variation between 2-5 days.

The duration between infection and case confirmation, 1/g

We set 1/*g*=8 days as a baseline. Ng. et al., and a previous work our ours estimated the duration from symptom onset to case confirmation in Singapore and China, respectively[3,5]. All works found that this duration decreases over time, likely because of awareness of the virus and thus behavior changes in seeking medical care. We took the lower bound estimates of the duration from symptom onset to case confirmation in Ng. et al. and our study with a range between 2-4 days, because by the time the virus caused major outbreaks in Europe, the virus is well known. If we assume an incubation period of 4-6 days. We get a range between 6-10 days for the duration from infection to case confirmation.

Distribution of duration from onset of infectiousness to death

We assumed an Erlang distribution for the duration from onset of infectiousness to death. As shown in Eqn. 4 in the main text, a realistic distribution is important for accurate parameter estimation. In the equation, $\left(\frac{nd}{r+nd}\right)^n$ are very different between n=1 and n>1. That is to say simply assuming an exponential distribution (i.e. n=1) lead to erroneous conclusions.

We and other showed that the mean time from symptom onset to death is between 16.5 and 18.5 days[3,6,7]. We estimated a shape parameter of this distribution to be between 4 and 5[3]. Therefore, we set 1/d=18.5 days for the mean duration between onset of infectiousness and death with ranges between 16.5 and 20.5 days.

Infection fatality ratio, X

We set X=0.01 as baseline with ranges between 0.04-0.015. This point estimates and range are consistent with several recent studies[7,8].

4. Parameter estimation

We fit the daily case confirmation function $\Psi(t)$ and the death count function $\Phi(t)$ to incidence data and daily death data to infer $\theta(t)$, r and I_0^*. Other parameter values are fixed according to previous estimates (see Table 1). The error is calculated as the residual sum of squares (RSS) between data and model predictions of incidence or death counts on a log scale. To compare between models, we calculate the Akaike Information Criterion (AIC) score as:

$$AIC = 2m_{par} + m_{data}\log(RSS/m_{data})$$

where m_{par} is the number of fitted parameters and m_{data} is the number of data points used in estimation.

5. Uncertainty quantification

To fully evaluate uncertainty in the estimated parameters given the dataset, we took a two-step approach. First, evaluate uncertainties in r, $\theta(t)$ and I_0^* while keeping other parameters fixed. We take a statistical sampling approach, sampling 10^6 parameter combinations by drawing parameters randomly from uniform distributions over the ranges specified in Table 1 and for the detection rate θ, we draw random numbers between 0·001 to 1 on a log scale. We accept parameter combinations within 2 ΔAIC scores of the lowest AIC score using the best-fit parameter combinations for each country. Second, to assess how estimation is impacted by uncertainties in variations in the fixed parameters, we used upper and lower bounds of the fixed parameter values (as shown in Table 1), and performed parameter estimation and assessed uncertainty in the estimated parameter values as described in the first step. Using only the upper and lower bounds is justified because the fixed parameters affect the values of the $\Psi(t)$ and $\Phi(t)$ functions monotonically. Therefore, using the extreme values would give the upper and lower bounds of the estimated parameter values. The upper and lower bounds reported in Fig. 1 and 2 were based on simulation results using all accepted parameter combinations in the two steps.

Supplementary References:

1. Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020.
2. Lauer SA, Grantz KH, Bi Q, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med* 2020.
3. Sanche S, Lin Y, Xu C, Romero-Severson E, Hengartner N, Ke R. The novel coronavirus, SARA-nCoV-2, is highly contagious and more infectious than initially estimated. . *Emerging Infectious Diseases* 2020; **(accepted)**.
4. Zou L, Ruan F, Huang M, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med* 2020.
5. Ng Y, Li Z, Chua YX, et al. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore — January 2–February 29, 2020. *MMWR Morb Mortal Wkly Rep* 2020.
6. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020; **395**(10229): 1054-62.
7. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infetious Diseases* 2020.
8. Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine* 2020.

Table S1. Dates of incidence data used for inference in the main text (15 days) and sensitivity analysis showing the estimated growth rate and detection rate are robust to choices of number of data points.

Country	Dates of incidence data used for inference (15 days)	Growth rate, r			Detection rate, θ		
		15 days of data points	13 days of data points	10 days of data points	15 days of data points	13 days of data points	10 days of data points
Belgium	Mar. 6 – 21	0.20	0.19	0.19	0.04	0.04	0.04
France	Feb. 29 – Mar. 14	0.22	0.24	0.22	0.03	0.03	0.03
Germany	Mar. 1 – 15	0.24	0.24	0.24	0.23	0.22	0.22
Italy	Feb. 23 – Mar. 8	0.24	0.26	0.28	0.02	0.02	0.02
Netherlands	Mar. 6 – 21	0.19	0.20	0.20	0.03	0.03	0.03
Spain	Mar. 2 – 16	0.29	0.28	0.33	0.02	0.02	0.01
Switzerland	Mar. 5 – 20	0.19	0.17	0.19	0.09	0.08	0.08
UK	Mar. 5 – 20	0.20	0.21	0.18	0.02	0.02	0.02
US	Mar. 3 – 17	0.28	0.27	0.28	0.04	0.04	0.04

Table S2. Comparison of models for the detection rate, θ, using AIC scores. The three models correspond to the three models listed in the Methods in the main text. Model 1, i.e. constant θ, is the best model for all 9 countries considered (bold AIC scores).

Country	Model 1 Constant	Model 2 Hill-type function	Model 3 Linear function
Belgium	**-45.6**	-39.6	-39.6
France	**-35**	-29.1	-29
Germany	**-43.6**	-37.7	-37.6
Italy,	**-58.6**	-52.6	-52.6
Netherlands	**-75.4**	-69.4	-69.4
Spain	**-54.8**	-49.1	-48.8
Switzerland	**-39.1**	-33.4	-33.1
UK	**-52.9**	-46.9	-46.9
US	**-40.9**	-35.1	-34.9